Project Proposal

Project Proposal: Predictive Modeling of Life Expectancy and Health Factors

1. Introduction: This project aims to build a predictive model for life expectancy using various health and socioeconomic factors. The dataset contains life expectancy, adult mortality, infant deaths, alcohol consumption, healthcare expenditure, vaccination coverage, HIV/AIDS prevalence, GDP, population, and other indicators for multiple countries over several years. We analyze this data to identify critical life expectancy factors and develop a robust predictive model.

2. Objectives:

- Explore the relationship between life expectancy and various predictor variables (Correlation matrix and pair plots to check if multicollinearity)
- Identify significant predictors of life expectancy through statistical analysis (Partial F- Test).
- Build a predictive model for life expectancy based on selected variables. (Multi Linear Regression)
- Evaluate the performance of the model and assess its predictive accuracy. (Check R^2_{adj}, Error plots)
- Apply transformations if necessary (heteroscedasticity appears)
- Verify the error plots again after all the transformations.

3. Deliverables:

- Finalized dataset with cleaned and preprocessed variables.
- A report documenting the findings of exploratory data analysis.
- Developed a predictive model for life expectancy with documented code.
- Evaluation metrics and insights regarding the model's performance.
- Recommendations based on model results.